

Partiel MAO Probabilités et Statistiques (18 mai 2021, 9h30-11h30)

Lorsque l'énoncé vous demande de **coder**, vous avez le choix entre :

1. Écrire du code Python. Dans ce cas on pourra supposer qu'on a appelé les modules habituels :

```
import numpy as np
import scipy as sp
import matplotlib.pyplot as plt
import numpy.random as rnd
import scipy.stats as sts
```

Vous ne perdrez pas de point si vous vous trompez dans les noms des fonctions et leurs options.

2. Écrire du pseudo-code, par exemple :

```

FONCTION comptage(p)
  n = 0
  POUR k de 1 à 10
    u = Uniforme([0, pi])
    SI u < p
      n = n+1
  RENVOYER n
```

Exercice 1.— Simulation de lois réelles

La loi de l'arcsinus sur $] -1, 1[$ est la loi de densité

$$f(x) = \frac{1}{\pi\sqrt{1-x^2}} \mathbb{1}_{]-1,1[}(x).$$

La loi du semi-cercle sur $[-2, 2]$ est la loi de densité

$$g(x) = \frac{1}{2\pi} \sqrt{4-x^2} \mathbb{1}_{[-2,2]}(x).$$

On souhaite simuler ces deux lois.

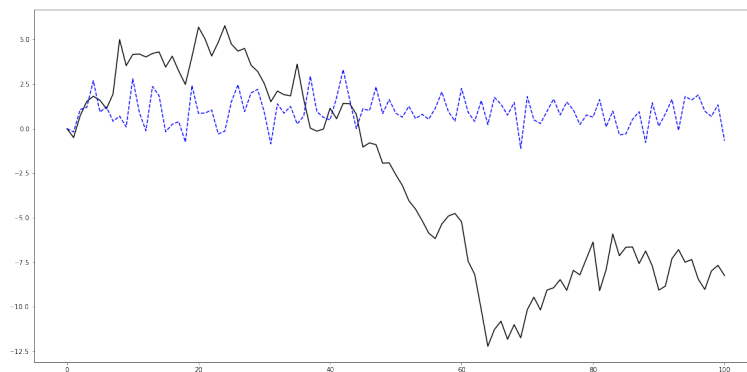
1. Expliquer comment simuler la loi de l'arcsinus par la méthode d'inversion. On rappelle qu'une primitive de la fonction $\frac{1}{\sqrt{1-x^2}}$ est la fonction $\arcsin(x)$.
Coder une fonction qui renvoie une réalisation de la loi de l'arcsinus.
2. On suppose qu'on a utilisé l'algorithme de la question 1 pour simuler N réalisations indépendantes, (X_1, \dots, X_N) . Proposer deux manières de visualiser le fait que cet échantillon suit bien la loi demandée (on ne demande pas de les coder).
3. Tracer à la main le graphe de la fonction g , ainsi que la densité de la loi uniforme sur $[-2, 2]$. Supposons qu'on sait simuler une variable uniforme sur $[0, 1]$, expliquer comment simuler la loi du semi-cercle par la méthode de rejet.
Coder une fonction qui renvoie une réalisation de la loi du semi-cercle.
4. Quelle est la loi du nombre d'essais dans l'algorithme de la question 3 ?

Exercice 2.— Une suite autorégressive

Soit a, b deux nombres réels, et soit $(Z_n)_{n \geq 1}$ une suite i.i.d. de variables aléatoires de loi normale centrée réduite $\mathcal{N}(0, 1)$. On considère la suite de variables aléatoires $(X_n)_{n \geq 0}$ définie par

$$X_0 = 0, \quad \text{et} \quad \forall n \in \mathbb{N}, \quad X_{n+1} = aX_n + b + Z_{n+1}.$$

1. Sur la figure 1, on a représenté deux réalisations de (X_0, \dots, X_{100}) . L'une des deux correspond à $a = 0.99$, $b = 0.01$, l'autre à $a = 0.01$, $b = 0.99$. À votre avis, laquelle correspond au premier choix, et laquelle correspond au second choix ?

FIGURE 1 – Deux réalisations de (X_0, \dots, X_{100}) .

2. On se demande si (X_n) converge vers une variable aléatoire X . Quel mode de convergence peut-on vraisemblablement exclure à partir de la figure 1 ?
3. On veut écrire du code qui réalise la figure 1. On propose :

```
def X(n,a,b):
    x = 0
    for i in range(n-1):
        x = a*x + b + sts.norm.rvs()
    return x

listeX1 = [X(n,0.99,0.01) for n in range(101)]
listeX2 = [X(n,0.01,0.99) for n in range(101)]

plt.plot(listeX1)
plt.plot(listeX2)
plt.show()
```

Qu'en pensez-vous ?

4. Montrer que pour tout $n \geq 1$, $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$, où μ_n, σ_n^2 sont à déterminer.
5. Pour quelles valeurs de a, b la suite (X_n) converge-t-elle en loi ? Dans ces cas-là, préciser sa limite.
Indication : on rappelle la fonction caractéristique d'une variable aléatoire $Z \sim \mathcal{N}(\mu, \sigma^2)$:
 $\phi_Z(t) = \mathbb{E}[e^{itZ}] = \exp\left(\mu it - \frac{\sigma^2 t^2}{2}\right)$.
On rappelle aussi le théorème de Lévy :
Une suite de v.a. réelles (Y_n) converge en loi vers Y ssi $\forall t \in \mathbb{R}, \phi_{Y_n}(t) \rightarrow \phi_Y(t)$.
6. Proposer une manière d'illustrer la convergence en loi précédente lorsqu'elle a lieu.
Coder votre méthode. On pourra utiliser librement la fonction X définie à la question 3.
7. On se demande maintenant pour quelles valeurs de a, b la suite (X_n) converge en probabilité. Pourquoi est-ce difficile à tester avec l'ordinateur ?
8. (*) Montrer que pour tout a, b , la suite (X_n) ne converge pas en probabilité.

Exercice 3.— Méthode de Monte-Carlo et estimation de l'erreur

On souhaite estimer l'intégrale

$$I = \int_0^1 g(x) dx$$

où g est une fonction mesurable bornée connue. Pour cela on cherche une suite iid de variables aléatoires $(X_n)_{n \geq 1}$ à support borné, telle que $\mathbb{E}[X_1] = I$. On note $S_n = \sum_{k=1}^n X_k$.

1. Donner un exemple de variables (X_n) vérifiant ces hypothèses et facilement simulables.
2. On voudrait que $\frac{S_n}{n}$ soit une approximation de I à ϵ près. Rappeler brièvement pourquoi on doit prendre n d'ordre au moins ϵ^{-2} .

3. Plus précisément, on se donne un $\epsilon > 0$ et un $\alpha \in]0, 1[$. On cherche un n tel que

$$\mathbb{P} \left(\left| \frac{S_n}{n} - I \right| > \epsilon \right) \leq \alpha. \quad (\star)$$

En utilisant Bienaymé-Tchebychev, donner un tel n en fonction de ϵ, α et de la fonction g .

4. On admet le résultat suivant, appelé *lemme de Hoeffding* :

Soit Y une variable aléatoire à support dans $[a, b]$ et centrée ($\mathbb{E}[Y] = 0$), alors pour tout $s > 0$,

$$\mathbb{E}[e^{sY}] \leq \exp \left(\frac{s^2(b-a)^2}{8} \right).$$

- (a) En utilisant le lemme de Hoeffding, montrer que pour tout $s > 0$,

$$\mathbb{P} \left(\frac{S_n}{n} - I > \epsilon \right) \leq \exp(-ns\epsilon + Cns^2)$$

où C est une constante à déterminer qui ne dépend que de la fonction g .

- (b) En déduire que

$$\mathbb{P} \left(\frac{S_n}{n} - I > \epsilon \right) \leq \exp \left(\frac{-n\epsilon^2}{4C} \right).$$

- (c) Montrer alors que

$$\mathbb{P} \left(\left| \frac{S_n}{n} - I \right| > \epsilon \right) \leq 2 \exp \left(\frac{-n\epsilon^2}{4C} \right).$$

- (d) En déduire une nouvelle valeur de n qui satisfait (\star) , et comparer avec le résultat de la question 2.